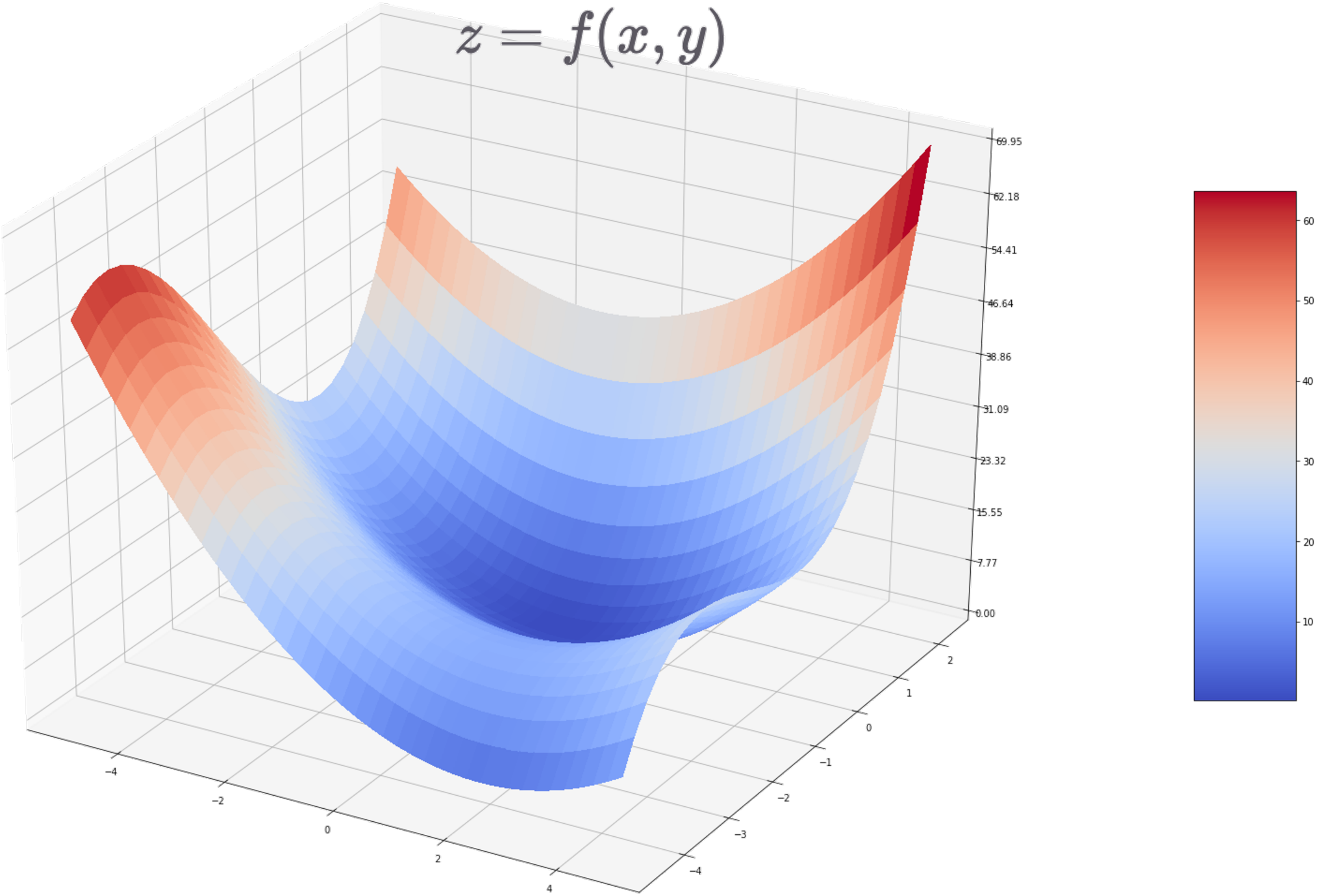
Appendix: Partial Derivative

Ki Hyun Kim

nlp.with.deep.learning@gmail.com

다변수 함수

- 여러 개의 변수를 입력으로 받는 함수



편미분

- 다변수 x와 y를 입력으로 받는 함수 f를 x로 미분할 경우,
  - 하나의 변수(아래 예시에서는 x)만 남겨놓고 나머지를 상수 취급하는 미분 방법

- 함수 f를 x변수(or 축)으로 미분
  - 편미분 기호 ∂

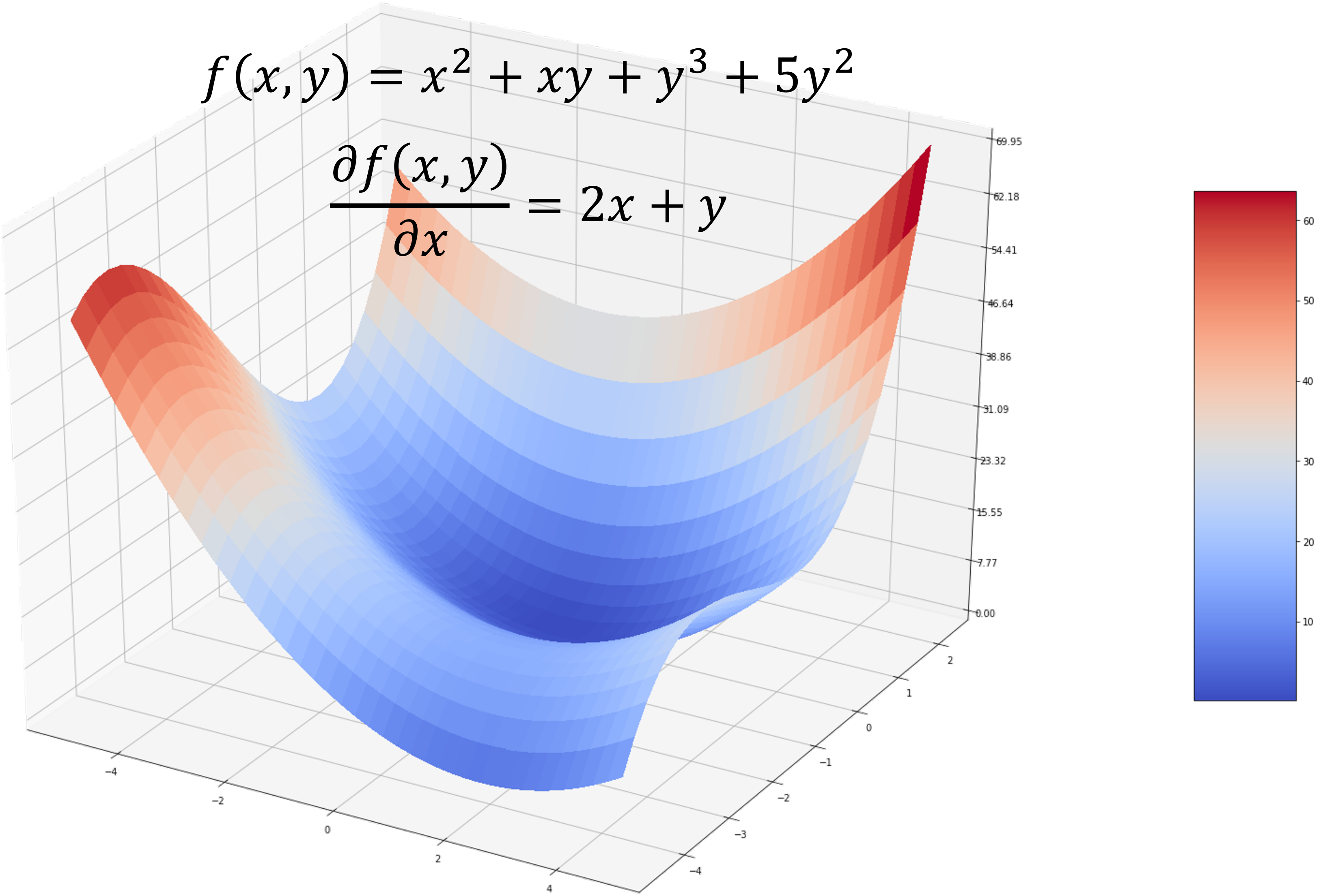
$$\frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{(x+h) - x}$$

편미분

- y값에 대해 뚝 잘랐을 때, x축에 대한 기울기

$$f(x,y) = x^2 + xy + y^3 + 5y^2$$

$$\frac{\partial f(x,y)}{\partial x} = 2x + y$$



함수의 입출력 형태

- 함수의 입력이 벡터인 경우

$$y = f(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}) = f(x),$$
$$\text{where } x \in \mathbb{R}^n.$$

- 함수의 출력이 벡터인 경우

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = f(x),$$
$$\text{where } y \in \mathbb{R}^n.$$

- 함수의 입력이 행렬인 경우

$$y = f(\begin{bmatrix} x_{1,1} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,m} \end{bmatrix}) = f(X),$$
$$\text{where } X \in \mathbb{R}^{n \times m}.$$

- 함수의 출력이 행렬인 경우

$$Y = \begin{bmatrix} y_{1,1} & \cdots & y_{1,m} \\ \vdots & \ddots & \vdots \\ y_{n,1} & \cdots & y_{n,m} \end{bmatrix} = f(x),$$
$$\text{where } Y \in \mathbb{R}^{n \times m}.$$

함수의 입출력 형태

- 입력과 출력이 벡터인 함수

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = f(x) = f(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}),$$
$$\text{where } f : \mathbb{R}^n \to \mathbb{R}^m.$$

스칼라를 벡터로, 스칼라를 행렬로 미분

- 미분 결과는 gradient 벡터가 되어 방향과 크기를 모두 나타냄

$$\frac{\partial f}{\partial x} = \nabla_x f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}, \quad \text{where } x \in \mathbb{R}^n.$$

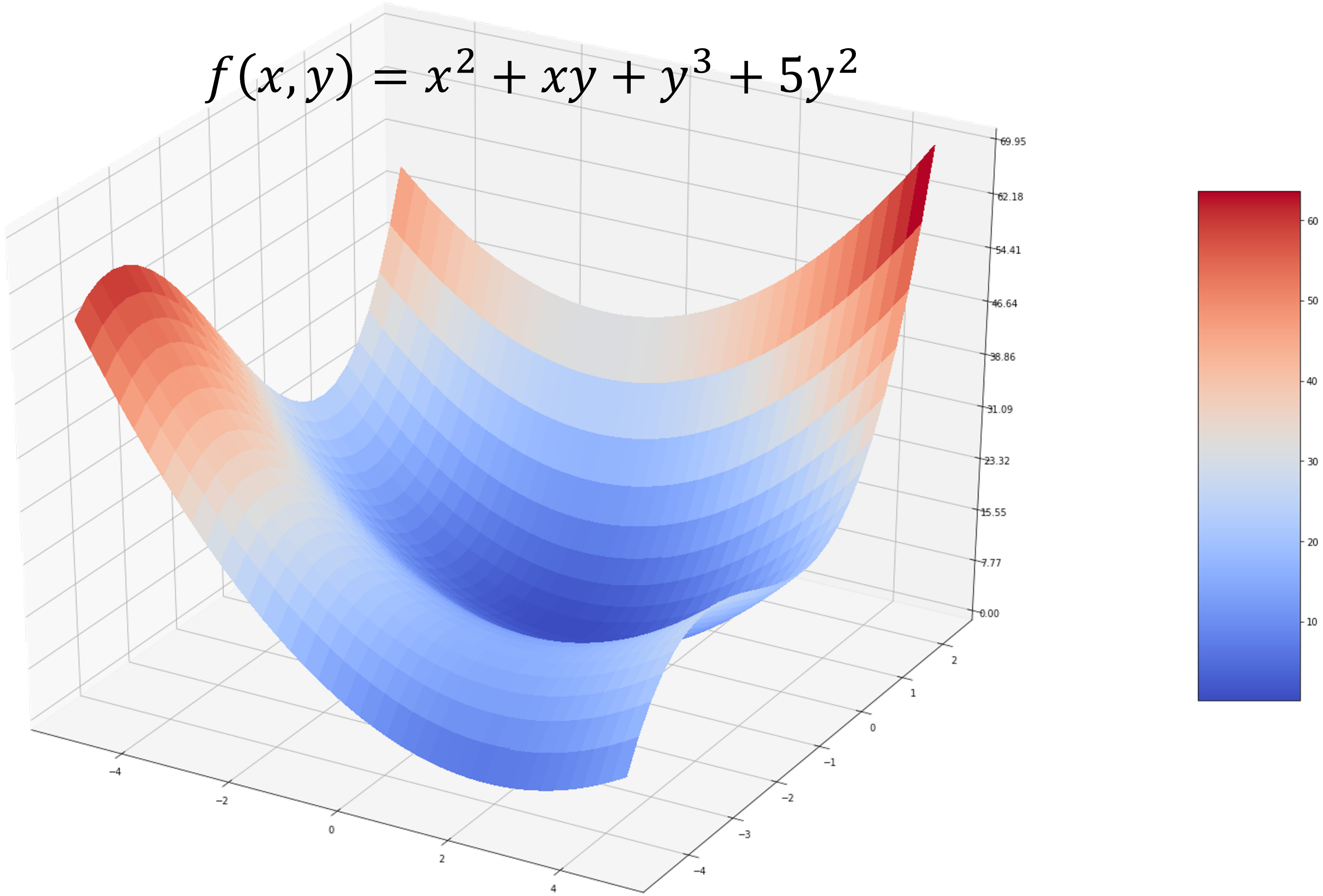
$$\frac{\partial f}{\partial X} = \nabla_X f = \begin{bmatrix} \frac{\partial f}{\partial x_{1,1}} & \cdots & \frac{\partial f}{\partial x_{1,m}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_{n,1}} & \cdots & \frac{\partial f}{\partial x_{n,m}} \end{bmatrix}, \quad \text{where } X \in \mathbb{R}^{n \times m}$$

Gradient

- 상미분과 달리 미분 결과가 벡터

$$\nabla f(x,y) = \begin{bmatrix} 2x + y \\ x + 3y^2 + 10y \end{bmatrix}$$

$$f(x,y) = x^2 + xy + y^3 + 5y^2$$



벡터를 스칼라로, 벡터를 벡터로 미분

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \cdots & \frac{\partial f_n}{\partial x} \end{bmatrix},$$
$$\text{where } f(x) \in \mathbb{R}^n.$$

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \cdots & \frac{\partial f_m}{\partial x} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix},$$
$$\text{where } x \in \mathbb{R}^n \text{ and } f(x) \in \mathbb{R}^m.$$

Why we learn this?

- Loss 함수 결과값 스칼라를 파라미터 행렬(θ)로 미분해야 한다면?

- DNN의 중간 결과물 벡터(h)를 파라미터 행렬(θ)로 미분해야 한다면?